



## Patch-based generation of a pseudo CT from conventional MRI sequences for MRI-only radiotherapy of the brain

Andreasen, Daniel; Van Leemput, Koen; Hansen, Rasmus H.; Andersen, Jon A. L.; Edmund, Jens Morgenthaler

*Published in:*  
Medical Physics

*Link to article, DOI:*  
[10.1118/1.4914158](https://doi.org/10.1118/1.4914158)

*Publication date:*  
2015

*Document Version*  
Publisher's PDF, also known as Version of record

[Link back to DTU Orbit](#)

*Citation (APA):*  
Andreasen, D., Van Leemput, K., Hansen, R. H., Andersen, J. A. L., & Edmund, J. M. (2015). Patch-based generation of a pseudo CT from conventional MRI sequences for MRI-only radiotherapy of the brain. *Medical Physics*, 42(4), 1596-1605. <https://doi.org/10.1118/1.4914158>

---

### General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

## Patch-based generation of a pseudo CT from conventional MRI sequences for MRI-only radiotherapy of the brain

Daniel Andreasen, Koen Van Leemput, Rasmus H. Hansen, Jon A. L. Andersen, and Jens M. Edmund

Citation: *Medical Physics* **42**, 1596 (2015); doi: 10.1118/1.4914158

View online: <http://dx.doi.org/10.1118/1.4914158>

View Table of Contents: <http://scitation.aip.org/content/aapm/journal/medphys/42/4?ver=pdfcov>

Published by the [American Association of Physicists in Medicine](#)

### Articles you may be interested in

[MRI-based treatment planning with pseudo CT generated through atlas registration](#)

Med. Phys. **41**, 051711 (2014); 10.1118/1.4873315

[A dual model HU conversion from MRI intensity values within and outside of bone segment for MRI-based radiotherapy treatment planning of prostate cancer](#)

Med. Phys. **41**, 011704 (2014); 10.1118/1.4842575

[Inter-slice bidirectional registration-based segmentation of the prostate gland in MR and CT image sequences](#)

Med. Phys. **40**, 123503 (2013); 10.1118/1.4829511

[Description and assessment of a registration-based approach to include bones for attenuation correction of whole-body PET/MRI](#)

Med. Phys. **40**, 082509 (2013); 10.1118/1.4816301

[CT substitute derived from MRI sequences with ultrashort echo time](#)

Med. Phys. **38**, 2708 (2011); 10.1118/1.3578928



**MULTI-CRITERIA OPTIMIZATION WILL CHANGE THE WAY YOU PLAN**  
Watch video >>

**THE NEW WAY**  
RayStation  
Balance clinical trade-offs with rayNavigator  
**PLAN APPROVED!**

**THE OLD WAY**  
If you change here, it affects the whole process  
Are you sure this was the best plan?  
Show the physician a better option?  
Acceptable dose to other OARs?  
Parotid glands saved?  
Respiration?  
Scan count and wait planning?

ADVANCING CANCER TREATMENT

RaySearch Laboratories

# Patch-based generation of a pseudo CT from conventional MRI sequences for MRI-only radiotherapy of the brain

Daniel Andreassen<sup>a)</sup>

*Department of Applied Mathematics and Computer Science, Technical University of Denmark, Kgs. Lyngby 2800, Denmark and Department of Oncology, Radiotherapy Research Unit, Copenhagen University Hospital, Herlev 2730, Denmark*

Koen Van Leemput

*Department of Applied Mathematics and Computer Science, Technical University of Denmark, Kgs. Lyngby 2800, Denmark and A.A. Martinos Center for Biomedical Imaging, MGH, Harvard Medical School, Charlestown, Massachusetts 02129*

Rasmus H. Hansen

*Department of Radiology, Copenhagen University Hospital, Herlev 2730, Denmark*

Jon A. L. Andersen and Jens M. Edmund

*Department of Oncology, Radiotherapy Research Unit, Copenhagen University Hospital, Herlev 2730, Denmark*

(Received 27 October 2014; revised 31 December 2014; accepted for publication 22 January 2015; published 17 March 2015)

**Purpose:** In radiotherapy (RT) based on magnetic resonance imaging (MRI) as the only modality, the information on electron density must be derived from the MRI scan by creating a so-called pseudo computed tomography (pCT). This is a nontrivial task, since the voxel-intensities in an MRI scan are not uniquely related to electron density. To solve the task, voxel-based or atlas-based models have typically been used. The voxel-based models require a specialized dual ultrashort echo time MRI sequence for bone visualization and the atlas-based models require deformable registrations of conventional MRI scans. In this study, we investigate the potential of a patch-based method for creating a pCT based on conventional  $T_1$ -weighted MRI scans without using deformable registrations. We compare this method against two state-of-the-art methods within the voxel-based and atlas-based categories.

**Methods:** The data consisted of CT and MRI scans of five cranial RT patients. To compare the performance of the different methods, a nested cross validation was done to find optimal model parameters for all the methods. Voxel-wise and geometric evaluations of the pCTs were done. Furthermore, a radiologic evaluation based on water equivalent path lengths was carried out, comparing the upper hemisphere of the head in the pCT and the real CT. Finally, the dosimetric accuracy was tested and compared for a photon treatment plan.

**Results:** The pCTs produced with the patch-based method had the best voxel-wise, geometric, and radiologic agreement with the real CT, closely followed by the atlas-based method. In terms of the dosimetric accuracy, the patch-based method had average deviations of less than 0.5% in measures related to target coverage.

**Conclusions:** We showed that a patch-based method could generate an accurate pCT based on conventional  $T_1$ -weighted MRI sequences and without deformable registrations. In our evaluations, the method performed better than existing voxel-based and atlas-based methods and showed a promising potential for RT of the brain based only on MRI. © 2015 American Association of Physicists in Medicine. [<http://dx.doi.org/10.1118/1.4914158>]

Key words: radiotherapy, magnetic resonance imaging, pseudo CT, MRI-only, patches

## 1. INTRODUCTION

Magnetic resonance imaging (MRI) is increasingly being used in modern radiotherapy (RT) treatment planning in combination with computed tomography (CT). The main reason is the superior soft tissue contrast which improves target and organ at risk (OAR) definition in the brain and other sites as compared to CT-based delineations.<sup>1-3</sup> In order to transfer the MRI delineations to the CT, the two scans must be aligned. This is usually done by a manual and/or automatic rigid registration between the CT and MRI scan.

It has been estimated that a mean cranial registration error of approximately 2 mm is introduced in this process<sup>4,5</sup> with similar values for other body sites such as the prostate.<sup>6</sup> These errors produce a systematic shift in the delineations and may ultimately lead to target under-dosage or an increased dose to the adjacent OARs.

Basing the entire RT chain on MRI as the only modality, so-called MRI-only RT, would remove these systematic errors and reduce patient discomfort as well as lower the workload and financial cost. It is, however, nontrivial to exclude the CT, since the MRI images do not contain information

about electron density which is needed for accurate dose calculations. Furthermore, with conventional MRI sequences, the signal from cortical bone is weak or nonexistent due to its rapid transversal signal relaxation time ( $T_2$ ) in the range 0.5–2 ms.<sup>7,8</sup> This means that MRI voxels containing bone are indistinguishable from air, and that patient setup based on digitally reconstructed radiographs (DRRs) is unfeasible. In order to solve these problems, electron density assignment must be done based on the MRI, which can then be used for dose calculation and patient setup in an RT setting.

Two different approaches for automatic density assignment have typically been taken to derive a so-called pseudo CT (pCT) from MRI: voxel-based and atlas-based. In the voxel-based approach, a pCT is generated from individual voxel intensities in the MRI scan, either by segmentation and subsequent bulk density assignment<sup>8–11</sup> or using a regression model to predict a continuous-valued pCT.<sup>11–14</sup> Common for these methods is the need for a specialized dual ultrashort echo time (dUTE) MRI sequence that captures the signal from components with a short  $T_2$  relaxation time. This makes bone voxels separable from air in the resulting MRI images.<sup>15</sup> However, adding this nonstandard sequence is not only inconvenient but also causes prediction errors in the pCT scans at tissue interfaces such as in the ear and nasal cavities. This has been attributed to partial volume and susceptibility effects as well as a low signal-to-noise ratio (SNR) of the dUTE scans.<sup>8,16,17</sup>

Atlas-based methods, on the other hand, estimate pCTs using conventional (non-dUTE) MRI sequences, and must therefore compensate for the bone/air ambiguity in the MRI images. This is achieved using deformable registration of one or multiple atlases of paired MRI/CT scans to the patient MRI scan and then using the warped atlas CT scan(s) as a pCT estimate.<sup>18–20</sup> The pure atlas-based methods rely on a correct deformable registration between atlas and patient MRI. This can be both time consuming (in the case of multiple atlases) and problematic if the patient is dissimilar to the atlas.<sup>8</sup> To correct for deformation errors, the method can be extended with a subsequent step of pattern recognition using Gaussian process regression based on 2D patches.<sup>21</sup> Though this improves robustness, the gain may be marginal compared to the added computational complexity.<sup>20</sup>

In the field of automated brain MRI segmentation, patch-based segmentation methods have recently been proposed with promising results.<sup>22,23</sup> In these approaches, 3D patches (i.e., small cuboidal image subregions) are extracted from the MRI and a spatially local search for the most similar patches in a preacquired database of labeled MRI scans is performed. The known labels of the resulting database patches are then fused to give the predicted label at each position. To facilitate the spatially local patch search, most patch-based methods use a rough linear alignment between the database MRI scans and the MRI to be segmented. The need for an accurate deformable registration is thus removed and the segmentation is driven mainly by patch similarities.

Based on these results, in this study, we investigate the potential of patch-based methods for predicting a continuous-valued pCT of the brain based on  $T_1$ -weighted MRI scans. In

particular, we incorporate the CT numbers as label information in the patch model to enable patch-based regression. We compare this method with two state-of-the-art methods in voxel-based and atlas-based pCT prediction, namely, (1) Gaussian mixture regression (GMR) based on dUTE scans<sup>12</sup> and (2) multiatlas information propagation based on  $T_1$ -weighted scans.<sup>19</sup> We perform a voxel-wise, geometric, and radiologic evaluation as well as a dosimetric evaluation for a photon treatment plan.

## 2. MATERIALS AND METHODS

### 2.A. Imaging

The data used in this study consisted of MRI and CT scans of five whole brain RT patients, three male and two female, aged 55–82 yr. The study was approved by the Capital Regional Ethics Committee (protocol number H-3-2011-107). The CT scans were acquired on a Philips Brilliance Big Bore CT with a voxel resolution  $0.6 \times 0.6 \times 2$  mm,  $512 \times 512 \times 110$  voxels using a standard protocol for brain scans (120 kV, 300 mAs). The MRI scans were acquired on a Philips Panorama 1 T open scanner. The sequences were a  $T_1$ -weighted 3D fast field echo (FFE), TE/TR = 6.9/25 ms, voxel resolution  $0.85 \times 0.85 \times 1.2$  mm,  $188 \times 188 \times 152$  voxels, and two dUTE scans at flip angles  $10^\circ$  and  $25^\circ$  with TE<sub>1</sub>/TE<sub>2</sub>/TR = 0.09/3.5/7.1 ms and an isotropic voxel resolution of 1 mm with  $256^3$  voxels. The scan time of a dUTE scan at one flip angle was approximately 7 min. The patients were fixed in treatment position during both the MRI and CT scanning using thermoplastic masks. Informed consent was obtained from all patients prior to acquiring the MRI scans additional to their standard imaging. Each patient's MRI/CT pair was rigidly aligned using mutual information with the default settings as implemented in 3D Slicer.<sup>24</sup> The CT scans were resliced and cropped to match the resolution and field of view of the MRI scans, thus generating dUTE/CT scan pairs at the dUTE scan resolution and  $T_1$ -weighted/CT scan pairs at the  $T_1$ -weighted scan resolution.

### 2.B. Patch-based pCT prediction

In order to facilitate an intensity-based similarity search, the tissues should have a consistent intensity throughout all  $T_1$ -weighted scans. To achieve this, a histogram-matching approach was used with two linear mapping segments separated at the average median intensity.<sup>25</sup>

A patch,  $P(x)$ , was defined as a cube with side length  $m$  voxels centered on the spatial location  $x$  in a MRI image. Similarly, a target value,  $T(x)$ , was defined for each  $P(x)$  as the Hounsfield unit (HU) value at  $x$  in the corresponding rigidly aligned CT image. For  $S$  patients, corresponding patches and target values were extracted for all positions,  $x$ , to create a database of patches,  $P_s(x)$ , with corresponding target values,  $T_s(x)$ , where  $s$  denotes one of  $S$  patients. Using this database, a pCT for a test patient was predicted by extracting patches from his/her MRI scan and doing an intensity-based nearest neighbor search in the patch database. For a patch at position  $y$

in the MRI of the test patient, the similarity measure was defined as

$$d(s,x) = \|P(y) - P_s(x)\|_2^2, \quad (1)$$

where  $d(s,x)$  denotes the squared  $L_2$ -norm between  $P(y)$  and  $P_s(x)$ . The search then consisted of finding the database patch that minimized  $d(s,x)$  and storing the corresponding patient and spatial position

$$(s^{\min}, x^{\min}) = \underset{s,x}{\operatorname{argmin}} d(s,x). \quad (2)$$

The search was extended to find the  $K$  most similar database patches and the  $K$  relevant target values were extracted from the database as  $T_{s_k^{\min}}(x_k^{\min})$  with  $k = 1, \dots, K$ . To assign a pCT HU value, a similarity-weighted average was computed

$$\text{pCT}(y) = \frac{\sum_k w_k \times T_{s_k^{\min}}(x_k^{\min})}{\sum_k w_k}, \quad (3)$$

with weights defined as

$$w_k = \exp\left(\frac{-d(s_k^{\min}, x_k^{\min})}{\min_k d(s_k^{\min}, x_k^{\min})}\right). \quad (4)$$

This ensured that if one patch was more similar than the rest, this patch would be weighted highly; conversely, if all patches were equally similar, they would be weighted equally.

To ensure that the local similarity search was indeed local and to limit the number of similarity comparisons required per patch, we applied the search space reduction and patch preselection method as described by Coupé *et al.*<sup>22</sup> An affine alignment of each database MRI and the patient MRI was performed and the search was constrained to only contain database patches in a local cuboidal search volume,  $V_{\text{search}}$ , of side length  $v$  voxels around  $y$ , such that  $x \in V_{\text{search}}$  in Eq. (2). The structural similarity measure (SSIM)<sup>26</sup> was used to discard highly dissimilar patches in the local neighborhood. This is a computationally simple similarity measure based on the mean and variance in each patch. All patches with SSIM  $< 0.95$  were discarded prior to the patch search performed in Eq. (2). If all patches were discarded in a search volume, the pCT value in the affected voxel was flagged as unknown. In postprocessing, such voxels were assigned the average pCT value of the closest assigned voxels. For the intensity fusion in Eq. (3), the  $K = 8$  most similar patches were used unless the SSIM only allowed a smaller number. An illustration of the patch-based pCT generation is shown in Fig. 1 for a simplified 2D case – the actual algorithm works in 3D.

## 2.C. Comparison of algorithms: GMR and multiatlas

The patch-based method was compared with Gaussian mixture regression based on dUTE scans<sup>12</sup> and multiatlas information propagation based on  $T_1$ -weighted scans;<sup>19</sup> two state-of-the-art techniques for pCT prediction within the voxel-based and atlas-based categories, respectively.

The multiatlas method consists of deformable registrations of multiple atlases of MRI/CT pairs to the patient MRI. A multiresolution  $B$ -spline transform in Elastix<sup>27</sup> was used for this purpose. Mutual information with 32 histogram bins was used as a metric and the final control point spacing was set to 5 mm. The local normalized cross correlation (LNCC) between the patient MRI and the atlas MRI was used to determine a voxel-wise ranking of each atlas CT before finally fusing the HU values. A high LNCC means a high degree of local similarity between the patient MRI and the deformed atlas MRI and thus results in a high ranking. Two parameters were involved in these steps, namely, a Gaussian kernel width,  $\sigma_{gk}$ , controlling the local neighborhood size involved in the LNCC calculation and a weight constant,  $\beta$ , controlling the decay of an exponential function used when converting the LNCC ranking to a fusing weight.

For the GMR method, the joint distribution of MRI and CT intensities was estimated as a mixture of multivariate Gaussian distributions using expectation maximization (EM).<sup>28</sup> Using this model, the expected value of the CT conditioned on newly observed MRI values can be found and used as a pCT prediction. Input images in addition to the dUTE scans were created using mean and standard deviation filters on the dUTE scans. A mask was then created to exclude air surrounding the head from the model training data. Fifty independent runs of  $k$ -means clustering with  $k$ -means++ initialization<sup>29</sup> were performed on the training data to estimate the initial values of the means, covariance matrices, and mixing proportions. The values resulting in the minimum total energy were chosen as initialization for the EM algorithm.

## 2.D. Model optimization

In order to ensure a fair comparison between the different methods, a nested cross validation scheme was used to find the optimal parameters for all methods. This meant that for each test patient, leave-one-out cross validation (LOOCV) was performed cyclically on the remaining four patients. In each of these four LOOCV folds, three patients were used to predict pCTs of the fourth using all possible combinations of a manually defined subset of model parameters. For each parameter configuration, the voxel-wise mean absolute error [MAE<sub>vox</sub>, as defined later in Eq. (5)] was used to evaluate the pCT. The parameter configuration that resulted in the average best performance across the four folds was then chosen as the optimal for that test patient and was used to predict his/her pCT using the remaining four patients. This scheme is well suited for model comparisons since each test patient is not involved in the parameter optimization, which gives an almost unbiased estimate of the true prediction error.<sup>30</sup> For the patch-based method, the optimal patch side length was found to be  $m = 5$  or  $m = 7$  voxels (depending on the patient) and the optimal search volume side length was  $v = 15$  voxels, yielding a maximum of  $15^3 \times 4 = 13\,500$  database patches to search through for each pCT voxel. Note that the value of  $K$ , used in Eq. (3), was set empirically to  $K = 8$  and therefore not optimized. For the multiatlas method, parameters for the LNCC ranking scheme were optimized. The Gaussian kernel

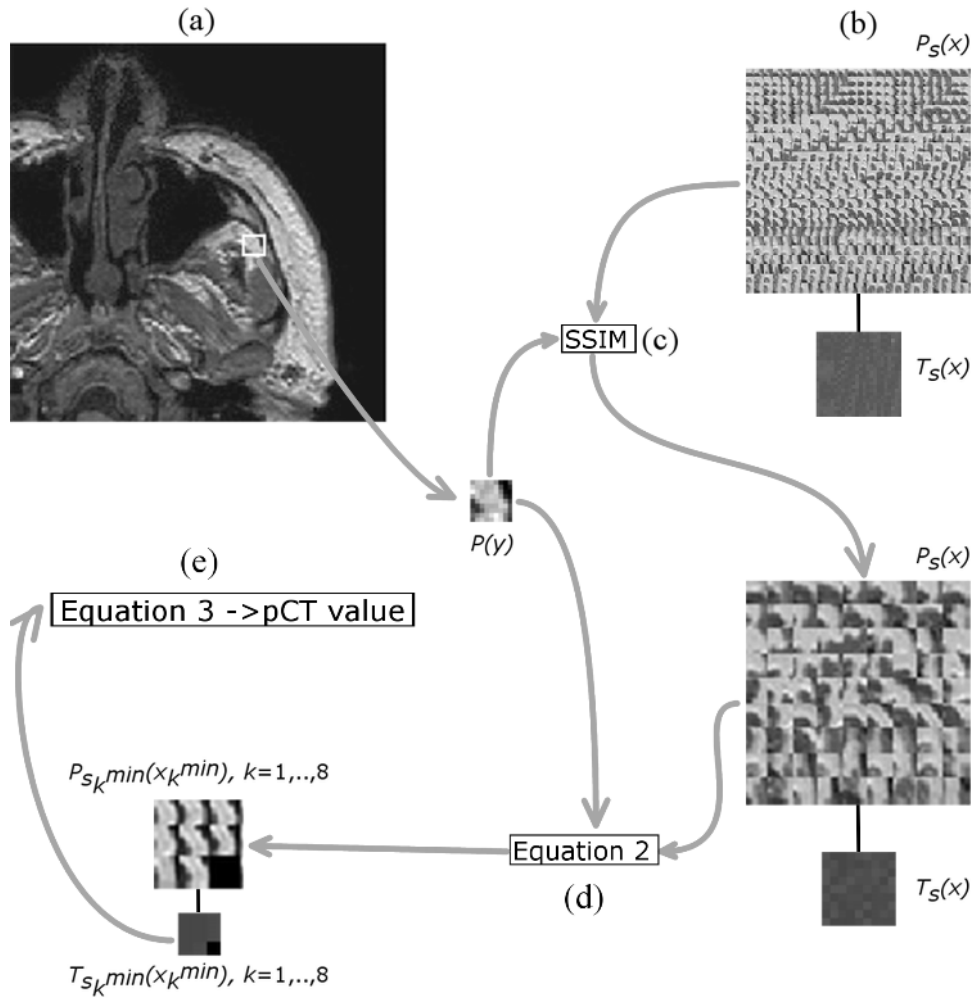


FIG. 1. 2D patch-based pCT prediction. (a) A patch,  $P(y)$ , is extracted from the test MRI at position  $y$ . (b) A database of spatially close patches,  $P_S(x)$ , stored with their corresponding CT value,  $T_S(x)$ . (c) Using the SSIM, highly dissimilar patches are discarded from the database. (d) Using Eq. (2), the eight most similar database patches,  $P_{S_k^{\min}}(x_k^{\min})$ , and their CT values,  $T_{S_k^{\min}}(x_k^{\min})$ , are found from the remaining database. (e) Using Eq. (3), the CT values,  $T_{S_k^{\min}}(x_k^{\min})$ , are combined to produce the final pCT value at position  $y$ .

width,  $\sigma_{gk}$ , was 5–9 voxels depending on the patient and the weight constant was  $\beta = 0.9$  for all patients. Note that the deformable registration parameters were chosen empirically and thus were not included in the nested cross validation. For the GMR method, the number of Gaussians to use in the model was found and set to 20–23, again depending on the patient.

**2.E. Geometric evaluation**

The pCTs were compared in terms of the voxel-wise mean absolute error  $MAE_{\text{vox}}$  in the head region (excluding surrounding air),

$$MAE_{\text{vox}} = \frac{1}{N} \sum_{i=1}^N |CT(i) - pCT(i)|, \tag{5}$$

where  $N$  is the total number of voxels in the head region. To reveal in which tissue regions errors were present, the  $MAE_{\text{vox}}$  in bins of 20 HU across the HU scale was also calculated. To determine whether the predictions were biased toward an

underestimation or overestimation of the real CT number, the mean voxel-wise error was calculated

$$ME_{\text{vox}} = \frac{1}{N} \sum_{i=1}^N [CT(i) - pCT(i)]. \tag{6}$$

To evaluate the correctness of the pCT bone geometry, the Dice similarity coefficient (DSC)<sup>31</sup> of bone was calculated

$$DSC = \frac{2(V_{\text{bone}}^{\text{pCT}} \cap V_{\text{bone}}^{\text{CT}})}{V_{\text{bone}}^{\text{pCT}} + V_{\text{bone}}^{\text{CT}}}, \tag{7}$$

where  $V_{\text{bone}}^{\text{pCT}}$  and  $V_{\text{bone}}^{\text{CT}}$  are the volumes of bone in the pCT and the real CT, respectively. We defined the bone volume as all voxels with a value > 200 HU giving a density of 1.14 g/cm<sup>3</sup> according to our CT calibration curve. DSC = 1 means complete overlap between the volumes and DSC = 0 means no overlap.

The  $MAE_{\text{vox}}$ ,  $ME_{\text{vox}}$ , and DSC are pure voxel-wise and geometric measures of the pCT accuracy. To provide a radiologic error measure, an evaluation based on the water

equivalent path length (WEPL) was introduced<sup>10,32</sup>

$$l' = \sum_i \Delta l_i \times \rho_i, \quad (8)$$

where  $\Delta l_i$  is the physical path length of voxel  $i$  and  $\rho_i$  is a radiological scaling factor that depends on the type of radiation and tissue. For MeV photons, it is the electron density relative to water. The value of  $\rho_i$  was found from the voxel CT number using a verified standard calibration curve in the treatment planning system. To compare WEPLs in the CT and pCT, a common point in both was defined at the level of the nasal cavity centered in the head. A sphere with its center in that point was defined, covering the entire head. The WEPL was then calculated in radial spokes from the center point toward the edge of the sphere. When the spokes traversed the edge of the head, the WEPL calculation was terminated so the length of each spoke varied with the patient anatomy. The spokes were defined in spherical coordinates with the center point as origin and with an angular spacing of  $2\pi/60$  in both the polar and azimuthal angles. All polar angles between  $0.5\pi$  and  $1.5\pi$  were excluded from the calculation, resulting in a volume including only the upper hemisphere of the head, covering most of the brain (see Fig. 2). In total, this yielded  $60 \times 31 = 1860$  spokes. The tissue was sampled every  $\Delta l = 0.02$  mm along each spoke. The CT number was then found at each sample by trilinear interpolation. To measure the difference in WEPL between the CT and pCT, the mean absolute WEPL error ( $\text{MAE}_{\text{WEPL}}$ ) was defined as

$$\text{MAE}_{\text{WEPL}} = \frac{1}{L} \sum_{j=1}^L |l'_j{}^{\text{CT}} - l'_j{}^{\text{pCT}}|, \quad (9)$$

where  $L$  is the number of spokes and  $l'_j$  is the WEPL of the  $j$ th spoke. Similarly, the mean WEPL error ( $\text{ME}_{\text{WEPL}}$ ) was defined as

$$\text{ME}_{\text{WEPL}} = \frac{1}{L} \sum_{j=1}^L l'_j{}^{\text{CT}} - l'_j{}^{\text{pCT}}. \quad (10)$$

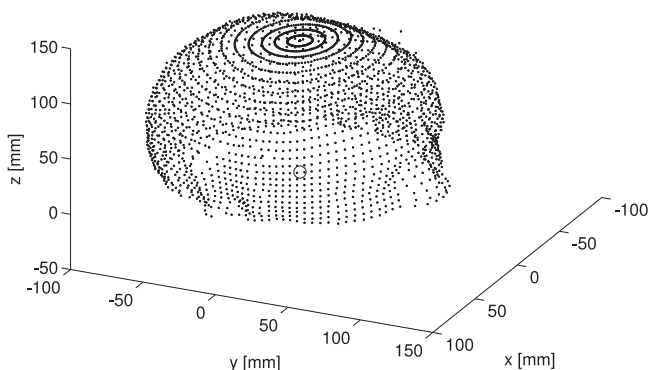


FIG. 2. The WEPL was calculated along spokes (not shown) from the center point (open circle) at  $(x, y, z) = (0, 0, 0)$  toward all points on the surface of the patient (dots). Along the spokes, the CT number was sampled and converted to relative electron density for WEPL calculation.

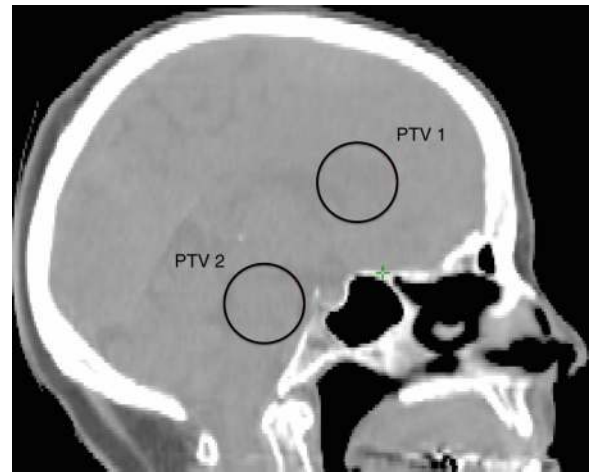


FIG. 3. Sagittal view of the two PTVs used in the dosimetric evaluation.

## 2.F. Dosimetric evaluation

A 6 MV photon treatment plan for two different spherical planning target volumes (PTVs) of 3 cm in diameter was created for all the patients. The first plan used a PTV positioned anterior to the center of the cerebrum (PTV 1, Fig. 3) with two  $15^\circ$  wedged lateral opposing fields and one anterior field, all equally weighted. The second plan had a PTV positioned behind the nasal cavity (PTV 2, Fig. 3) with four equally weighted fields, two lateral opposing, and two anterior/posterior opposing. For all fields, the apertures were cropped to the PTV plus a 0.5 cm margin. The PTVs were chosen to represent theoretically easy and difficult cases, respectively, with PTV 1 positioned in a homogeneous part of the brain and PTV 2 having a large degree of heterogeneity in the tissue composition of its surroundings. The treatment planning was carried out on each patient's pCT in Eclipse v11.0 (Varian Medical Systems, Inc., Palo Alto, CA), prescribing a mean dose of 2 Gy in 30 fractions ( $D_{\text{pre}} = 60$  Gy) to the PTV and calculating the dose distribution using the analytical anisotropic algorithm (AAA). The plan was then transferred to the real CT and recalculated using the same plan parameters and monitor units. The PTV dose volume histogram (DVH) was used to evaluate the dosimetric difference between CT and pCT. The percentage point deviation was calculated for the DVH points relevant to PTV coverage, i.e., the near-minimum ( $D_{98\%}$ ), near-maximum ( $D_{2\%}$ ), and median ( $D_{\text{median}}$ ) absorbed dose.<sup>33</sup> Furthermore, a normalized mean absolute dosimetric error ( $\text{nMAE}_{\text{dos}}$ ) was calculated as

$$\text{nMAE}_{\text{dos}} = \frac{1}{M} \sum_{i=1}^M \frac{|D_{\text{vol}(i)}^{\text{CT}} - D_{\text{vol}(i)}^{\text{pCT}}|}{D_{\text{pre}}}, \quad (11)$$

where  $M$  is the total number of DVH points between  $D_{100\%}$  and  $D_{0\%}$  in dose increments of 0.1 Gy and  $D_{\text{vol}(i)}$  is the accumulated dose in Gy given to  $\text{vol}(i)\%$  of the volume in either the CT or pCT. Similarly, the normalized mean dosimetric error ( $\text{nME}_{\text{dos}}$ ) was calculated as

$$\text{nME}_{\text{dos}} = \frac{1}{M} \sum_{i=1}^M \frac{D_{\text{vol}(i)}^{\text{CT}} - D_{\text{vol}(i)}^{\text{pCT}}}{D_{\text{pre}}}, \quad (12)$$

$\text{nMAE}_{\text{dos}}$  explains the magnitude of the errors and  $\text{nME}_{\text{dos}}$  reveals if the errors are biased toward underestimation or overestimation. The combined dosimetric score (CDS) is a combination which penalizes bias in the errors<sup>11</sup>

$$\text{CDS} = \text{nMAE}_{\text{dos}} + |\text{nME}_{\text{dos}}|. \quad (13)$$

### 3. RESULTS

In Fig. 4, transverse slices of the real CT and predicted pCTs are shown for the different methods. Visually, the results based on  $T_1$ -weighted MRI in (c) and (d) are better than the dUTE-based in (b), especially in the nasal cavities.

Figure 5 shows the  $\text{MAE}_{\text{vox}}$  calculated as a function of the real CT value in bins of 20 HU and averaged for the five patients. In general, GMR shows the highest errors in the bone region ( $> 200$  HU), fat region (approximately  $[-100; -50]$  HU), and the region between air ( $-1000$  HU) and fat tissue. This HU range is dominated by values of partial volume effects in the real CT. The patch-based and multiatlas methods have similar performance in most regions. Table I shows the average voxel-wise, geometric, and radiologic errors for the different methods. On average, the patch-based and multiatlas methods have the lowest  $\text{MAE}_{\text{vox}}$  and the highest DSC, with a slight favor for the patch-based method. Looking at the  $\text{ME}_{\text{vox}}$ , the patch-based and multiatlas methods have average values closer to 0, indicating no consistent pattern in the errors. However, given the magnitude of the standard deviation for GMR,  $\text{ME}_{\text{vox}} = 0$  is also a plausible value for this method. Looking at the WEPL evaluation in Table I, the patch-based and multiatlas methods have the smallest  $\text{MAE}_{\text{WEPL}}$ , with slightly lower values for the patch-based method.

In Fig. 6, the results of the dosimetric evaluation are shown. For all methods, the metrics show average smaller errors, i.e., values closer to 0, for PTV 1 compared to PTV 2, illustrating the more challenging position of PTV 2. The difference in performance between the methods is also generally smaller for PTV 1 than for PTV 2. Looking at PTV 1, the  $\text{nMAE}_{\text{dos}}$  is similar for the patch method and GMR, but the  $\text{nME}_{\text{dos}}$  reveals that GMR has a bias toward overestimating the dose. The multiatlas method has the largest  $\text{nMAE}_{\text{dos}}$ , with a bias ( $\text{nME}_{\text{dos}}$ ) toward underestimating the dose. Looking at the coverage of PTV 1 (lower panels), the patch-based method has an average deviation close to 0%. The other methods have deviations  $< 0.5\%$ . When looking at PTV 2, the average  $\text{nMAE}_{\text{dos}}$  is the lowest for the patch-based method, with GMR and multiatlas having higher values and larger standard deviation. We observe that both the GMR and multiatlas methods have a bias toward underestimating the dose as seen in the upper middle panel. With regards to the dose coverage (lower panels), the patch-based method has the smallest average deviations, except for  $D_{98\%}$ , where GMR has a slightly smaller deviation. On average, multiatlas has a better performance than GMR in  $D_{\text{median}}$  and  $D_{2\%}$ .

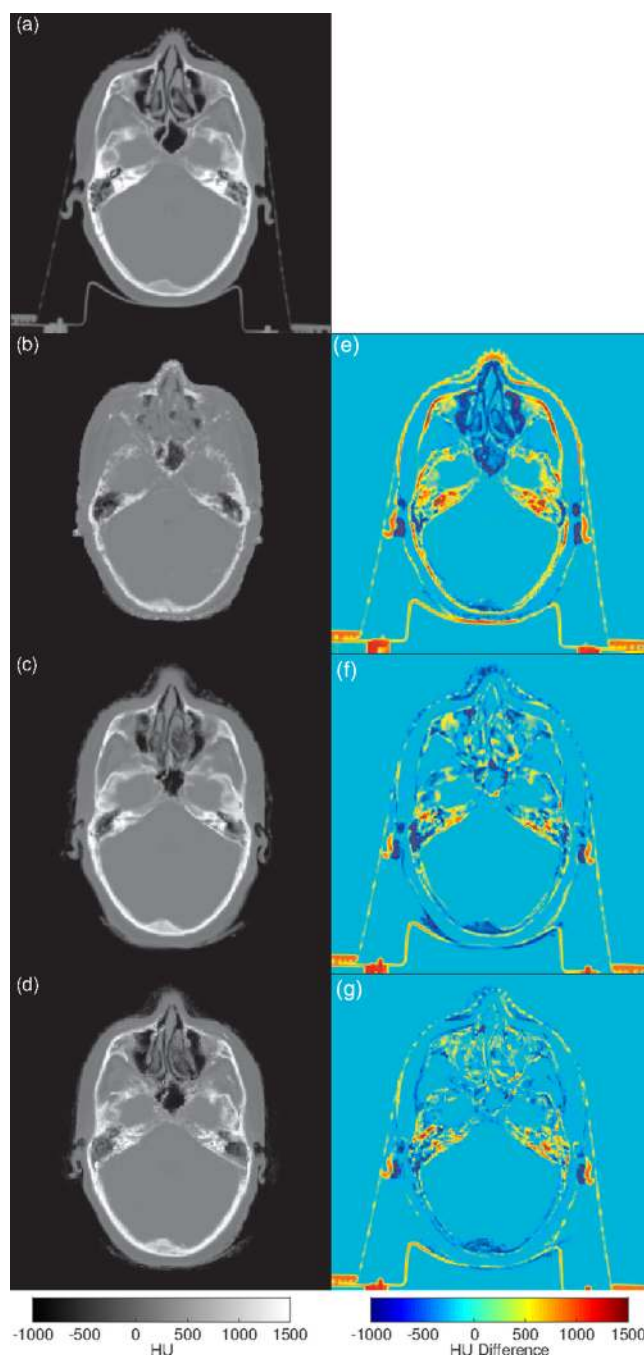


FIG. 4. Transverse slices for comparison of pCTs with real CT. (a)–(d) show the real CT, the GMR pCT, the multiatlas pCT, and the patch-based pCT, respectively. (e)–(g) show the difference maps between the real CT and the GMR, the multiatlas, and the patch-based pCTs, respectively. Negative values indicate an overestimation of the HU value and positive values indicate an underestimation.

### 4. DISCUSSION

In this paper, we evaluated a patch-based method for predicting brain pCTs based on conventional  $T_1$ -weighted MRI images. The method required no deformable registrations and was shown to yield comparable or better results than existing methods using Gaussian mixture regression on dUTE scans or multiatlas information propagation on  $T_1$ -weighted scans.



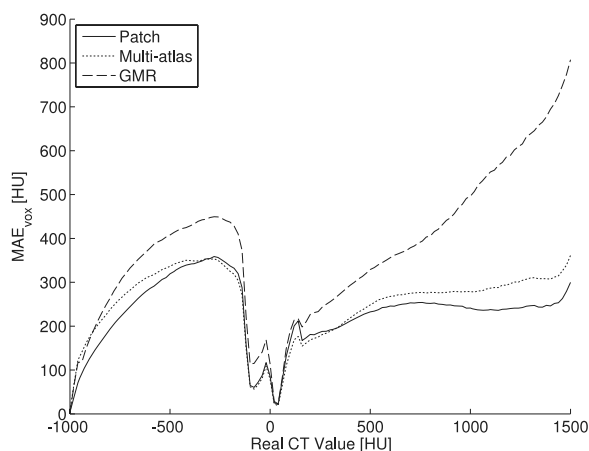


FIG. 5.  $MAE_{vox}$  calculated in bins of 20 HU and averaged across the five patients.

In terms of the  $MAE_{vox}$  curves shown in Fig. 5, the dUTE-based method showed the largest voxel-wise errors in most tissue regions of the brain. The errors we observed in the fat region could be caused by the data being recorded on a 1 T scanner where water/fat phase cancellation occurs approximately at 3.5 ms after excitation, i.e., at  $TE_2$  of the dUTE sequence. In a previous study, we tested if adding a Dixon sequence as GMR model input would aid in discriminating fat/water and bone.<sup>34</sup> Though this improved bone predictions, it did not improve predictions in fat voxels. Rank *et al.* reported errors in pCT bone prediction due to fatty-tissue appearing hyperintense in dUTE scans acquired at 3 T.<sup>10</sup> This suggests that issues in the fat region may be independent of field strength and not due to phase cancellations.

In the bone region, the two methods based on the  $T_1$ -weighted MRI had the smallest errors, even though bone is not uniquely defined in terms of intensity in these images. The deformable registration used in the multiatlas method can compensate for this lack of information, assuming a successful alignment of the atlas MRIs and the patient MRI. In the present study, this assumption seemed to hold, but as implied earlier, this may not always be the case if the anatomical variation is large. Furthermore, the deformable registration introduces another set of parameters to adjust, and it can be hard to

TABLE I. The voxel-wise and geometric quality measures: mean absolute voxel-wise error ( $MAE_{vox}$ ), mean voxel-wise error ( $ME_{vox}$ ) in HU, and DSC of bone volume. Radiologic measures: mean absolute WEPL error ( $MAE_{WEPL}$ ) and mean WEPL error ( $ME_{WEPL}$ ) of the water equivalent path lengths (in mm). Average value and standard deviation ( $\sigma$ ) for the five patients are shown.

	Patch-based	Multiatlas	GMR
$MAE_{vox}$ (HU)	85 ( $\sigma = 14$ )	97 ( $\sigma = 19$ )	148 ( $\sigma = 22$ )
$ME_{vox}$ (HU)	1 ( $\sigma = 14$ )	-4 ( $\sigma = 17$ )	22 ( $\sigma = 28$ )
DSC	0.84 ( $\sigma = 0.02$ )	0.83 ( $\sigma = 0.01$ )	0.67 ( $\sigma = 0.03$ )
$MAE_{WEPL}$ (mm)	2.2 ( $\sigma = 1.0$ )	2.7 ( $\sigma = 0.8$ )	4.8 ( $\sigma = 1.3$ )
$ME_{WEPL}$ (mm)	0.4 ( $\sigma = 1.8$ )	-0.6 ( $\sigma = 1.9$ )	1.1 ( $\sigma = 2.1$ )

find one configuration that provides a successful registration for all patients. Here, we tested a few parameter settings and chose the one that resulted in the average lowest  $MAE_{vox}$ . For a more unbiased estimate of the prediction performance of the multiatlas method, the registration parameters could be included in the nested cross-validation.

The patch-based method, on the other hand, achieved a lower error in the bone region without a deformable registration. Instead, the neighborhood information contained in each patch in combination with the constrained search volume ensured that bone and air patches were not confused.

In terms of the voxel-wise errors and DSC presented in Table I, the patch-based method had the best performance. Johansson *et al.* reported an average  $MAE_{vox}$  of 137 HU for five patients using the GMR method,<sup>12</sup> which is within the standard deviation of our GMR findings. The GMR results could potentially be improved by adding spatial information to the Gaussian mixture model or using different MRI reconstruction techniques.<sup>16,35</sup> This was not investigated further in our study. For the multiatlas method, Burgos *et al.* reported an average  $MAE_{vox}$  of 102 HU ( $\sigma = 10$ ) for seven patients,<sup>19</sup> which also agrees with our findings. We included the DSC measure since the bone geometry is important for generating DRRs. As was shown in Fig. 5, the patch and multiatlas methods have the highest prediction accuracy in the bone region, which is also reflected in the DSC results. This suggests that these methods are better suited for DRR generation. In a qualitative evaluation, Jonsson *et al.* described an overall acceptable quality of DRRs generated on the basis of GMR pCTs, except in the nasal cavity and sphenoidal sinuses.<sup>36</sup> From our results, it seems that the pCTs based on  $T_1$ -weighted MRI visually have a better quality in this region. However, the quality of the resulting DRRs and their potential for treatment setup was not investigated further in the present study. A factor that could influence the pCT predictions is the nonuniform intensity variations intrinsic to MRI scans. In the present study, we did not apply a correction for this. From a visual inspection, the acquired MRI scans showed only minor nonuniform intensity variations in the imaged volume, which is probably due to the relatively low field strength. For the patch-based method, applying a bias field correction did not improve predictions. It could potentially improve predictions of the other methods, however, and it will be an important step for the patch-based method at higher field strengths and/or larger fields of view.

The WEPL evaluation was introduced to provide a radiologically more relevant error measure than the pure voxel-wise errors. It takes into account the piece-wise linear relationship between the HU and electron density and also provides a simplistic imitation of the treatment simulation where radiation encounters several tissues when sent through the patient. The errors in WEPL are dependent on the position of the center point and the direction of the radial spokes. Here, we chose to evaluate the whole upper hemisphere of the head to produce an average estimate for all possible planning scenarios. Maybe, due to this averaging, the results of the WEPL evaluation did not diverge from the trends observed in the voxel-wise and geometric evaluation.

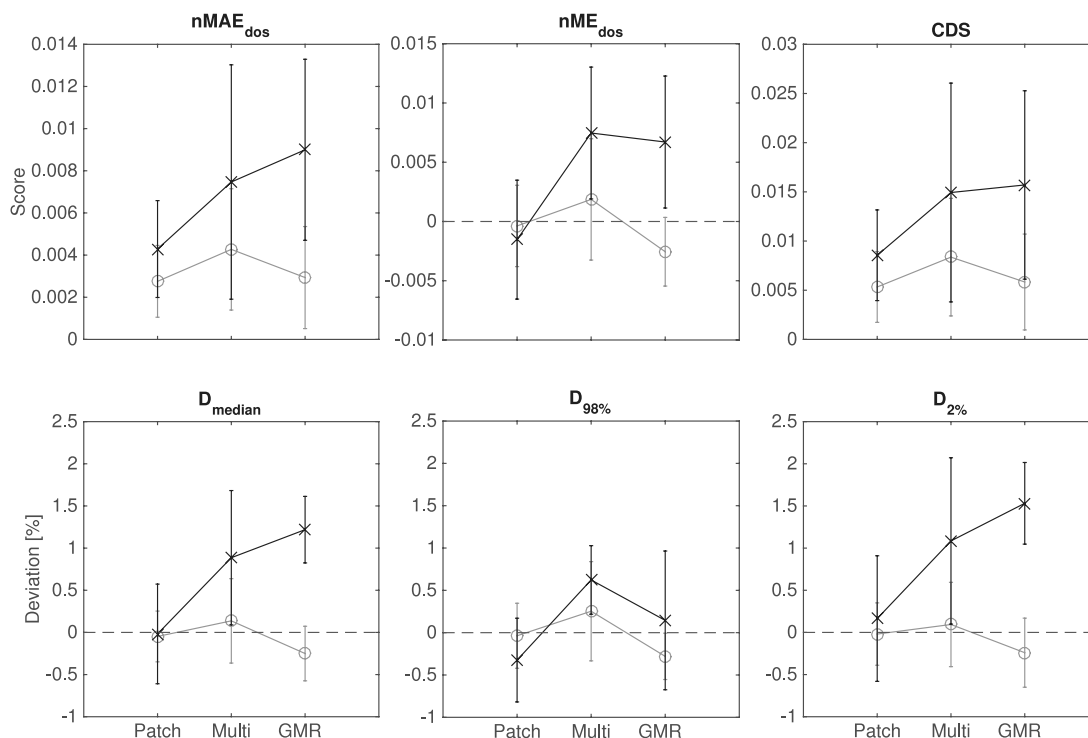


Fig. 6. Dosimetric errors for the two PTVs. PTV 1: gray lines and open circles. PTV 2: black lines and crosses. Dashed line indicates zero. Top row: normalized mean absolute dosimetric error ( $nMAE_{dos}$ ), normalized mean dosimetric error ( $nME_{dos}$ ), and CDS. Bottom row: percentage point deviation in  $D_{median}$ ,  $D_{98\%}$ , and  $D_{2\%}$ . Average values are shown along with  $\pm\sigma$  interval.

For the dosimetric evaluation, an easy and a challenging PTV site was chosen, which was reflected in the magnitude of the errors in Fig. 6. Overall, the patch-based method had the lowest average errors, following the trend from the other evaluations. Looking at the CDS, GMR performed well and was on par with the patch-based method for PTV 1 and with multiatlas for PTV 2. We obtained a comparable value of around 0.02 for the CDS of GMR for a target similar to PTV 2 in a previous study.<sup>11</sup> It should be noted that since the CDS only accounts for the magnitude of  $nME_{dos}$ , the direction of the error is removed. When averaging across patients, this means that a method with consistent overestimation of the dose can get the same CDS as a method where the direction of error is random. This effect explains why the patch method and GMR have similar CDS for PTV 1, even though the patch method performs better in  $nME_{dos}$ .

In  $D_{98\%}$ , GMR performed well for PTV 2 but with a larger standard deviation. In a dosimetric evaluation of the GMR method, Jonsson *et al.* reported a percentage point deviation of 0.86% in  $D_{90\%}$  (Ref. 36) for a target somewhat similar to PTV 2 in the present study. Though  $D_{90\%}$  is less sensitive to changes in the DVH shape, it seems that the dosimetric error is of the same order as our findings in  $D_{98\%}$ .

A statistical criterion for the reliable use of MRI-only states that for 95% of the patients, the maximum uncertainty in the DVH points related to target coverage should be within 2%.<sup>37</sup> Although a study with more patients is still needed, we note that for PTV 1, all methods fulfilled this criterion. This speaks in favor of a reliable use of MRI-only for targets positioned away from cavities. We also note that the patch-based method

fulfilled the criterion for the challenging PTV 2, showing potential for a more general use of MRI-only RT.

From the voxel-wise and radiologic evaluation, we expected that the multiatlas method would have dosimetric errors closer to those of the patch-based method. This did not seem to be the case, especially for PTV 2, where the multiatlas and GMR methods were closer in performance. Therefore, even though the voxel-wise and radiologic error may serve as a proxy for the dosimetric performance, they should be accompanied by a dosimetric evaluation to get the full picture. This agrees with our previous findings.<sup>11</sup>

Comparing the two methods based on  $T_1$ -weighted MRI, the voxel-wise and radiologic differences were small with a slight favor of the patch-based method. In terms of dosimetry, the patch-based method had an average better performance, especially in PTV 2. The advantage of both methods is that they work on any MRI sequence as long as an atlas or database of MRI/CT pairs has been obtained. This means that they can be adapted to the clinical practice without the need for extra sequences and scan time. The patch-based method further has the advantage that it relies only on linear registrations to provide a rough alignment of the database MRI and patient MRI. Linear registrations can be performed faster than deformable ones, which could potentially make the patch-based method faster than the multiatlas method. In the current Matlab implementation, however, using a brute force search for the most similar patches, it took approximately 15 hours to predict a pCT with four database patients. On the same hardware, the multiatlas and GMR methods took roughly 36.5 min (including deformable registrations) and

6.5 min, respectively. None of the implementations were optimized for speed, and especially the patch-based method could be improved substantially by using an approximate nearest neighbor (ANN) patch search algorithm such as OPAL.<sup>38</sup> A subsequent implementation of an ANN algorithm with similarities to OPAL, reduced the pCT prediction time for one patient to 38 min with an increase in MAE<sub>vox</sub> of 9 HU. We believe this can be further improved, but it was not the focus of the present study.

The patch-based method is still dependent on anatomical similarity, even without the deformable registration. Indeed, in smaller regions of the brain in one patient, we found that all patches in the database were discarded due to the SSIM being below the threshold value. These dissimilar regions could be present in the MRI of patients with tumors or other brain abnormalities that are not found in the patch database. This poses a problem for the method. However, based on the SSIM, we were able to produce a pCT with problematic (dissimilar) regions marked for quality assurance. This is in contrast to the multiatlas approach, where it can be hard to know where the deformable registrations were unsuccessful without manual inspection of each atlas. In the cases where no patches matched, we assigned the average pCT value of the neighboring assigned voxels, which did not seem to affect the accuracy of the pCT. Still, this way of handling the problem may prove too simple in the general case, especially if the dissimilar regions are large or if the tissue type of the neighboring voxels differs from that of the region. Increasing the size of the patient database could potentially alleviate some of this issue, but a focus of our future work is to incorporate a more advanced system for outlier handling.

As mentioned in the introduction, a pattern recognition step using Gaussian process regression (GPR) based on 2D patches has previously been investigated for improving an atlas-based method using deformable registrations.<sup>21</sup> The presented patch-based method shares some similarities with the GPR method since both are examples of so-called kernel smoothers. The difference lies in the way the kernel is defined, which is done explicitly in Eq. (4) for the patch-based method and implicitly through a covariance function in the GPR method. Furthermore, the patch-based method uses linear registrations, so the assumption of exact spatial correlation between the patient and database is not met. Therefore, the position of a patch does not affect its contribution weight, which is the case in the GPR method. Finally, the patch-based method adaptively normalizes the kernel for each test patch with the minimum  $L_2$ -norm in Eq. (4).

An average MAE<sub>vox</sub> of 100.7 HU was reported with the GPR method on  $T_1$ -weighted scans for 17 patients,<sup>21</sup> which is close to our results using the multiatlas or patch-based methods. It was not reported whether this value was calculated for the entire image volume or only for the head region, as in the present study.

Overall, the methods for pCT prediction based on similarities in conventional MRI scans seem promising for MRI-only RT. Interpatient anatomical variability does introduce an uncertainty in the pCTs, but the higher voxel-wise, geometric, and dosimetric accuracy compared to current voxel-based

methods may make this a worthwhile compromise. A larger study of the robustness of the patch-based method and the dosimetric uncertainty is still needed and is part of our future work.

## 5. CONCLUSION

In this study, we showed that a patch-based method could generate a pCT based on a conventional  $T_1$ -weighted MRI sequence without using deformable registrations or special dUTE sequences. We demonstrated a competitive performance of the method in several quality measures when compared to state-of-the-art atlas-based and voxel-based methods. In terms of dosimetric accuracy, the patch-based method showed a promising potential for use in MRI-only RT of the brain including PTVs positioned in challenging regions.

## ACKNOWLEDGMENT

This work was supported by a research grant from Varian Medical Systems, Inc.

<sup>a)</sup> Author to whom correspondence should be addressed. Electronic mail: dana@dtu.dk

<sup>1</sup>R. Prabhakar, K. Hareesh, T. Ganesh, R. Joshi, P. Julka, and G. Rath, "Comparison of computed tomography and magnetic resonance based target volume in brain tumors," *J. Cancer Res. Ther.* **3**(2), 121–123 (2007).

<sup>2</sup>M. Ahmed, M. Schmidt, A. Sohaib, C. Kong, K. Burke, C. Richardson, M. Usher, S. Brennan, A. Riddell, M. Davies, K. Newbold, K. J. Harrington, and C. M. Nutting, "The value of magnetic resonance imaging in target volume delineation of base of tongue tumours – A study using flexible surface coils," *Radiother. Oncol.* **94**(2), 161–167 (2010), Selected papers from the 10th Biennial ESTRO Conference on Physics and Radiation Technology for Clinical Radiotherapy.

<sup>3</sup>C. Rasch, R. Steenbakkers, and M. van Herk, "Target definition in prostate, head, and neck," *Semin. Radiat. Oncol.* **15**(3), 136–145 (2005).

<sup>4</sup>G. M. Cattaneo, M. Reni, G. Rizzo, P. Castellone, G. L. Ceresoli, C. Cozzarini, A. J. M. Ferreri, P. Passoni, and R. Calandrino, "Interobserver variability and impact of image registration of MR(pre-operative) images on treatment planning CT scans," *Radiother. Oncol.* **75**(2), 217–223 (2005).

<sup>5</sup>K. Ulin, M. M. Urie, and J. M. Cherlow, "Results of a multi-institutional benchmark test for cranial CT/MR image registration," *Int. J. Radiat. Oncol., Biol., Phys.* **77**(5), 1584–1589 (2010).

<sup>6</sup>T. Nyholm, M. Nyberg, M. Karlsson, and M. Karlsson, "Systematisation of spatial uncertainties for comparison between a MR and a CT-based radiotherapy workflow for prostate treatments," *Radiat. Oncol.* **4**, 54 (2009).

<sup>7</sup>I. L. Reichert, M. D. Robson, P. D. Gatehouse, T. He, K. E. Chappell, J. Holmes, S. Girgis, and G. M. Bydder, "Magnetic resonance imaging of cortical bone with ultrashort TE pulse sequences," *Magn. Reson. Imaging* **23**(5), 611–618 (2005).

<sup>8</sup>V. Keereman, Y. Fierens, T. Broux, Y. D. Deene, M. Lonnew, and S. Vandenberghe, "MRI-based attenuation correction for PET/MRI using ultrashort echo time sequences," *J. Nucl. Med.* **51**, 812–818 (2010).

<sup>9</sup>Y. Berker, J. Franke, A. Salomon, M. Palmowski, H. C. Donker, Y. Temur, F. M. Mottaghy, C. Kuhl, D. Izquierdo-Garcia, Z. A. Fayad, F. Kiessling, and V. Schulz, "MRI-based attenuation correction for hybrid PET/MRI systems: A 4-class tissue segmentation technique using a combined ultrashort-echo-time/dixon MRI sequence," *J. Nucl. Med.* **53**(5), 796–804 (2012).

<sup>10</sup>C. M. Rank, N. Hünemohr, A. M. Nagel, M. C. Röthke, O. Jäkel, and S. Greilich, "MRI-based simulation of treatment plans for ion radiotherapy in the brain region," *Radiother. Oncol.* **109**(3), 414–418 (2013).

<sup>11</sup>J. M. Edmund, H. M. Kjer, K. Van Leemput, R. H. Hansen, J. A. Andersen, and D. Andreassen, "A voxel-based investigation for MRI-only radiotherapy of the brain using ultra short echo times," *Phys. Med. Biol.* **59**(23), 7501 (2014).

- <sup>12</sup>A. Johansson, M. Karlsson, and T. Nyholm, "CT substitute derived from MR sequences with ultrashort echo time," *Med. Phys.* **38**(5), 2708–2714 (2011).
- <sup>13</sup>A. Johansson, M. Karlsson, J. Yu, T. Asklund, and T. Nyholm, "Voxel-wise uncertainty in CT substitute derived from MRI," *Med. Phys.* **39**(6), 3283–3290 (2012).
- <sup>14</sup>D. Andreassen, "Creating a pseudo-CT from MRI for MRI-only based radiation therapy planning," Master's thesis, Technical University of Denmark, DTU Compute, Denmark, 2013, E-mail: compute@compute.dtu.dk, Matematiktorvet, Building 303-B, DK-2800 Kgs. Lyngby.
- <sup>15</sup>J. Rahmer, U. Blume, and P. Börner, "Selective 3D ultrashort TE imaging: Comparison of 'dual-echo' acquisition and magnetization preparation for improving short-T2 contrast," *Magn. Reson. Mater. Phys., Biol. Med.* **20**(2), 83–92 (2007).
- <sup>16</sup>A. Johansson, A. Garpebring, M. Karlsson, T. Asklund, and T. Nyholm, "Improved quality of computed tomography substitute derived from magnetic resonance (MR) data by incorporation of spatial information – Potential application for MR-only radiotherapy and attenuation correction in positron emission tomography," *Acta Oncol.* **52**(7), 1369–1373 (2013).
- <sup>17</sup>C. Rank, C. Tremmel, N. Hünemohr, A. Nagel, O. Jäkel, and S. Greulich, "MRI-based treatment plan simulation and adaptation for ion radiotherapy using a classification-based approach," *Radiat. Oncol.* **8**(51), 1–13 (2013).
- <sup>18</sup>J. A. Dowling, J. Lambert, J. Parker, O. Salvado, J. Fripp, A. Capp, C. Wratten, J. W. Denham, and P. B. Greer, "An atlas-based electron density mapping method for magnetic resonance imaging (MRI)-alone treatment planning and adaptive MRI-based prostate radiation therapy," *Int. J. Radiat. Oncol., Biol., Phys.* **83**(1), e5–e11 (2012).
- <sup>19</sup>N. Burgos, M. Cardoso, M. Modat, S. Pedemonte, J. Dickson, A. Barnes, J. Duncan, D. Atkinson, S. Arridge, B. Hutton, and S. Ourselin, "Attenuation correction synthesis for hybrid PET-MR scanners," *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2013*, Lecture Notes in Computer Science Vol. 8149, edited by K. Mori, I. Sakuma, Y. Sato, C. Barillot, and N. Navab (Springer, Berlin, Heidelberg, 2013), pp. 147–154.
- <sup>20</sup>J. Uh, T. E. Merchant, Y. Li, X. Li, and C. Hua, "MRI-based treatment planning with pseudo CT generated through atlas registration," *Med. Phys.* **41**(5), 051711 (8pp.) (2014).
- <sup>21</sup>M. Hofmann, F. Steinke, V. Scheel, G. Charpiat, J. Farquhar, P. Aschoff, M. Brady, B. Schölkopf, and B. J. Pichler, "MRI-based attenuation correction for PET/MRI: A novel approach combining pattern recognition and atlas registration," *J. Nucl. Med.* **49**, 1875–1883 (2008).
- <sup>22</sup>P. Coupé, J. V. Manjón, V. Fonov, J. Pruessner, M. Robles, and D. L. Collins, "Patch-based segmentation using expert priors: Application to hippocampus and ventricle segmentation," *NeuroImage* **54**(2), 940–954 (2011).
- <sup>23</sup>F. Rousseau, P. Habas, and C. Studholme, "A supervised patch-based approach for human brain labeling," *IEEE Trans. Med. Imaging* **30**(10), 1852–1862 (2011).
- <sup>24</sup>A. Fedorov, R. Beichel, J. Kalpathy-Cramer, J. Finet, J. C. Fillion-Robin, S. Pujol, C. Bauer, D. Jennings, F. Fennessy, M. Sonka, J. Buatti, S. Aylward, J. Miller, S. Pieper, and R. Kikinis, "3D Slicer as an image computing platform for the quantitative imaging network," *Magn. Reson. Imaging* **30**(9), 1323–1341 (2012).
- <sup>25</sup>L. Nyul, J. Udupa, and X. Zhang, "New variants of a method of MRI scale standardization," *IEEE Trans. Med. Imaging* **19**(2), 143–150 (2000).
- <sup>26</sup>Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.* **13**(4), 600–612 (2004).
- <sup>27</sup>S. Klein, M. Staring, K. Murphy, M. A. Viergever, and J. P. W. Pluim, "Elastix: A toolbox for intensity-based medical image registration," *IEEE Trans. Med. Imaging* **29**(1), 196–205 (2010).
- <sup>28</sup>A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *J. R. Stat. Soc., Ser. B* **39**(1), 1–38 (1977).
- <sup>29</sup>D. Arthur and S. Vassilvitskii, "K-means++: The advantages of careful seeding," in *Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA '07* (Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, 2007), pp. 1027–1035.
- <sup>30</sup>S. Varma and R. Simon, "Bias in error estimation when using cross-validation for model selection," *BMC Bioinf.* **7**(1), 91 (2006).
- <sup>31</sup>L. R. Dice, "Measures of the amount of ecologic association between species," *Ecology* **26**(3), 297–302 (1945).
- <sup>32</sup>S. Mori, G. T. Chen, and M. Endo, "Effects of intrafractional motion on water equivalent pathlength in respiratory-gated heavy charged particle beam radiotherapy," *Int. J. Radiat. Oncol., Biol., Phys.* **69**(1), 308–317 (2007).
- <sup>33</sup>International Commission on Radiation Units and Measurements (ICRU), "Report 83: 3. Special considerations regarding absorbed-dose and dose-volume prescribing and reporting in IMRT," *J. ICRU* **10**(1), 27–40 (2010).
- <sup>34</sup>D. Andreassen, J. L. Andersen, R. H. Hansen, K. Van Leemput, and J. M. Edmund, "The impact of a Dixon sequence in creating a pseudo CT scan from MR images using a Gaussian mixture regression model," *Radiother. Oncol.* **106**(2), S229 (2013).
- <sup>35</sup>A. Johansson, A. Garpebring, T. Asklund, and T. Nyholm, "CT substitutes derived from MR images reconstructed with parallel imaging," *Med. Phys.* **41**(8), 082302 (7pp.) (2014).
- <sup>36</sup>J. H. Jonsson, A. Johansson, K. Söderström, T. Asklund, and T. Nyholm, "Treatment planning of intracranial targets on MRI derived substitute CT data," *Radiother. Oncol.* **108**(1), 118–122 (2013).
- <sup>37</sup>M. Korsholm, L. Waring, and J. Edmund, "A criterion for the reliable use of MRI-only radiotherapy," *Radiat. Oncol.* **9**(1), 16 (2014).
- <sup>38</sup>V. T. Ta, R. Giraud, D. Collins, and P. Coupé, "Optimized patchmatch for near real time and accurate label fusion," *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2014*, Lecture Notes in Computer Science edited by P. Golland, N. Hata, C. Barillot, J. Hornegger, and R. Howe (Springer International Publishing, 2014), Vol. 8675, pp. 105–112.